

**PLEASE NOTE THAT THIS APPLICATION IS
ASSOCIATED WITH A CUSTOMER NUMBER. PLEASE
INPUT THE CUSTOMER NUMBER LOCATED BENEATH
THE SIGNATURE LINE OF THE ATTORNEY.**

THANK YOU

Client's ref.: 91059/
File:0548-8323dus/final/spin(王琮郁)

TITLE

SPLIT GATE FLASH MEMORY CELL

5

CROSS REFERENCE TO RELATED APPLICATION

This is a divisional application of co-pending U.S. Patent Application Serial No.10/307,704, filed on December 2, 2002, which claims the benefit of priority to U.S. Provisional Application 60/383, 481, filed May 24, 2002.

10

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates in general to a semiconductor device and method for fabricating the same. More particularly, it relates to a reduced-size split gate flash memory cell that increases integration with ICs.

15

Description of the Related Art

Non-volatile memory, such as flash memory, stores data regardless of electrical power supplied, and reads and writes data by controlling a threshold voltage of a control gate. Conventionally, flash memory includes a floating gate and a control gate. The floating gate stores charge and the control gate reads and writes data. Since flash memory has a high operating speed, it is widely applied for consumer electrical goods, such as digital cameras, mobile phones, personal stereos, and laptops.

25

FIGs. 1A-1F are cross-sections showing a conventional method of fabricating a split gate flash memory cell.

First, in FIG. 1A, a silicon substrate 10 is provided, and a thin silicon oxide layer 12 is formed thereon serving as a tunneling oxide layer. The tunneling oxide layer 12 can be formed by thermal oxidation. Next, a polysilicon layer 14 and a silicon nitride layer 16 are sequentially deposited on the tunneling oxide layer 12.

Next, in FIG. 1B, a photoresist layer 18 is coated on the silicon nitride layer 16, leaving a portion exposed. Thereafter, the exposed portion of the silicon nitride layer 16 is etched to form an opening 20 exposing the polysilicon layer 14.

Next, in FIG. 1C, the photoresist layer 18 is stripped and thermal oxidation is performed on the exposed polysilicon layer 14 using the remaining silicon nitride layer 16a as a mask to form a thick oxide layer 24 having tipped and thin portions 24a, 24b at its edge.

Next, in FIG. 1D, with the remaining silicon nitride layer 16a is removed by wet etching to expose the polysilicon layer 14.

Next, in FIG. 1E, the polysilicon layer 14 is etched by anisotropic etching using thick oxide layer 24 as a mask to the tunneling oxide layer 12. The remaining polysilicon layer 14a is used as a floating gate.

Finally, in FIG. 1F, a gate dielectric layer 28, a control gate 30, and source region S/drain region D are formed and the tunneling oxide layer 12 uncovered by control gate 30 and the floating gate 14a is removed to finish the fabrication of the split gate flash memory cell.

However, in the fabrication of the conventional split gate flash memory, it is difficult to align the control gate and control the critical dimension (CD) of the control gate. Moreover, since increasing the integration of ICs is imperative, the conventional split gate flash memory is no longer suitable for small devices.

SUMMARY OF THE INVENTION

Accordingly, an object of the invention is to provide a novel split gate flash memory cell and method of fabricating the same to reduce the size of the flash memory cell by forming the source line and the floating gate in the trench of the substrate.

According to one aspect, the invention provides a method of fabricating a split gate flash memory cell. First, a substrate having a trench is provided, and a conductive stud insulated from the substrate is formed in the lower trench serving as a source line. Next, a source region is formed in the substrate adjacent to the upper conductive stud. Next, an insulating layer is formed on the conductive stud. Next, a conductive spacer is formed on the upper sidewall of the trench serving as a floating gate, protruding and insulated from the substrate. Next, an insulating stud is formed on the insulating layer, wherein the insulating stud is higher than the conductive spacer. Next, a first conductive layer is formed over the substrate of the outside conductive spacer, wherein the first conductive layer is insulated from the conductive spacer and the substrate, respectively. Next, a first insulating spacer is formed on the sidewall of the insulating stud to cover a part of the first conductive layer. Next, the first conductive layer is removed using the first insulating spacer as a mask to expose

the substrate and the remaining conductive layer is used as a control gate. Finally, a drain region is formed in the exposed substrate.

5 According to another aspect, the invention provides a split gate flash memory cell. The memory cell includes a substrate having a trench, a conductive stud disposed in the lower trench and insulated from the substrate serving as a source line, a source region formed in the substrate adjacent to the upper conductive stud, and an insulating layer disposed on the
10 conductive stud. A conductive spacer is disposed on the upper sidewall of the trench serving as a floating gate, protruding and insulated from the substrate. An insulating stud is disposed on the insulating layer, with the top thereof higher than that of the conductive spacer. A first conductive layer is disposed
15 over the substrate of the outside conductive spacer serving as a control gate, and is insulated from the conductive spacer and the substrate, respectively. A first insulating spacer disposed on the sidewall of the insulating stud covers the first conductive layer, and a drain region is formed in the substrate
20 of the outside first conductive layer.

The conductive spacer and the first conductive layer are doped polysilicon. The insulating layer is high density plasma oxide. The first insulating stud is silicon oxide or boron silicate glass. The first insulating spacer is silicon nitride.

25

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention can be more fully understood by reading the subsequent detailed description in conjunction with the examples and references made to the accompanying drawings,
30 wherein:

FIGS. 1A-1F are cross-sections showing a conventional method of fabricating a split gate flash memory cell.

FIGS. 2-19 are cross-sections showing a method of fabricating a split gate flash memory cell according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

A preferred embodiment of the present invention is now described with reference to FIGS. 2-19.

10 In FIG. 2, a substrate 200, such as a silicon wafer, is provided. Next, a first mask layer 201 and a second mask layer 205 are formed on the substrate 200 in sequence. In this invention, the first mask layer 201 is composed of a pad oxide 202 and a silicon nitride layer 204, and the second mask layer
15 205 is composed of a pad oxide 206 and a silicon nitride layer 208. Next, an opening is formed in the second and first mask layers 205, 201 to expose the substrate 200 by lithography and etching. Finally, the substrate 200 under the opening is etched to form a deep trench 212 having a predetermined depth, for
20 example, about 3~7 μm .

In FIG. 3, a conformable liner oxide 214 is formed on the second mask layer 205 and the surface of the trench 212 by conventional deposition, such as chemical vapor deposition (CVD). Thereafter, a conductive layer 216, such as polysilicon
25 layer, is deposited over the second mask layer 205 and filled into the trench 212. Next, the excess conductive layer 216 over the second mask layer 205 is removed by chemical mechanical polishing (CMP) or dry etching. Next, the liner oxide 214 and the conductive layer 216 in the trench 212 are etched back by
30 anisotropic etching, such as dry etching, or isotropic etching,

such as wet etching, to reduce their levels below the surface of the substrate 200. Next, a conformable insulating layer (not shown), such as silicon nitride layer, is formed on the second mask layer 205 and the surface of the trench 212 by CVD. Thereafter, the insulating layer is etched by anisotropic etching, such as dry etching, to form an insulating spacer 218 on the sidewall of the trench 212.

In FIG. 4, the conductive layer 216 in the trench 212 is etched back again to lower the conductive layer 216 below the insulating spacer 218. Subsequently, the liner oxide 214 between the insulating spacer 218 and the conductive layer 216 is removed by isotropic etching, such as wet etching, to expose a part of the substrate 200. The remaining conductive layer 216 in the lower trench 212 forms a conductive stud 216' serving as a source line insulated from the substrate 200 by the remaining liner oxide 214. Next, a conformable conductive layer 220, such as a doped polysilicon layer, is formed on the second mask layer 205 and the surface of the trench 212 by conventional deposition, such as CVD. Next, a doping region S is formed in the exposed substrate 200 of the trench 212 adjacent to the upper conductive stud (source line) 216' by high temperature drive-in, serving as a source region.

In FIG. 5, the conductive layer 220 over the conductive stud 216' is etched back to leave a part of conductive layer 220 on the liner oxide 214 to make connection between the conductive stud 216' and the doping region S. Thereafter, the insulating spacer 218 is removed, and a conformable insulating layer 222, such as high density plasma oxide (HDP oxide), is formed on the second mask layer 205 and the surface of the trench 212 by CVD. Here, the top and bottom of the insulating layer 222 are thicker

than the sidewall of the insulating layer 222. Next, the trench 212 is filled with a photoresist layer 224 which is etched back to below the top of the trench 212.

5 In FIG. 6, the insulating layer 224 on the sidewall of the trench 212 uncovered by the photoresist 224 is removed by isotropic etching, such as wet etching, using the photoresist layer 224 as a mask to expose the sidewall of the silicon nitride layer 208. Next, after the photoresist layer 224 is removed, the sidewall of the silicon nitride layer 208 is etched by
10 isotropic etching, such as wet etching, using the remaining insulating layer 222' as a mask to expose the pad oxide 206 (etching stop layer) and form an opening 225 having a step profile.

FIGS. 7-9 are cross-sections showing a method of forming
15 a floating gate in the trench 212.

In FIG. 7, the insulating layer 222' on the silicon nitride layer 208 and the sidewall of the trench 212 and the pad oxide 206 in the opening 225 are removed by isotropic etching, such as wet etching, to leave the insulating layer 222" on the
20 conductive stud 216' and expose the first mask layer 201 in the opening 225 and the substrate 200 in the trench 212, respectively. Next, an oxide layer 226 is formed on the exposed substrate 200 in the trench 212 by thermal oxidation. Next, a conductive layer 228, such as doped polysilicon, having a
25 thickness about 200~400 Å is formed on the second mask layer 205 and the surface of the opening 225. Thereafter, a sacrificial layer 230, such as photoresist, is formed on the second mask layer 205 and filled into the opening 225 and the trench 212. Next, a part of the sacrificial layer 230 is removed to lower
30 it below the top of the opening 225. Here, the top of the

remaining sacrificial layer 230 is higher than the conductive layer 228 formed on the bottom of the opening 225.

In FIG. 8, the conductive layer 228 uncovered by the remaining sacrificial layer 230 is removed by isotropic etching, using the remaining sacrificial layer 230 as a mask.

In FIG. 9, after the remaining sacrificial layer 230 is removed, the conductive layer 228 is etched back by anisotropic etching, such as dry etching, to expose the insulating layer 222" and form a conductive spacer 228' on the sidewall of the upper trench 212 protruding the top of the substrate 200. Here, the conductive spacer 228' having a tip portion is used as a floating gate, and the tip portion discharges during erasing. Moreover, the oxide layer 226 is used as a gate oxide. Next, an insulating layer 232, such as oxide, is filled into the opening 225 by conventional deposition, such as low-pressure CVD (LPCVD). Subsequently, the insulating layer 232 is lowered below the top of the second mask layer 205. Next, a cap layer 234, such as silicon oxynitride or polysilicon, is deposited on the second mask layer 205 and the insulating layer 232. Thereafter, the excess cap layer 232 on the second mask layer 205 is removed by CMP.

FIGS. 10 and 11 are cross-sections showing a method of forming an insulating stud on the insulating layer 222".

In FIG. 10, the silicon nitride layer 208 is etched using the cap layer 234 as a mask and the pad oxide 206 as an etching stop layer to expose a part of the insulating layer 232.

In FIG. 11, the exposed sidewall of the insulating layer 232 and the pad oxide 206 are etched by isotropic etching, such as wet etching, to expose the silicon nitride layer 204 and form

an insulating stud 232' over the insulating layer 222" protruding the conductive spacer 228'.

FIGS. 12-15 are cross-sections showing a method of forming a control gate on the substrate 200 of the outside floating gate 228'.

5 In FIG. 12, the cap layer 224 and the first mask layer 201 are sequentially removed to expose the substrate 200. Next, conformable oxide layer 236 having a thickness of about 100-500 Å and conductive layer 238, such as doped polysilicon, are sequentially formed on the surface of the substrate 200, the protruding floating gate 228', and the insulating stud 232'. Thereafter, a photoresist layer 240 is coated on the conductive layer 238. Next, the photoresist layer 240 is lowered below the top of the insulating stud 232' to expose a part of conductive layer 238 over the sidewall and top of the insulating stud 232'.

15 In FIG. 13, the exposed conductive layer 238 is etched using the photoresist layer 240 as a mask to leave the conductive layer 238 over the substrate 200 of the outside floating gate 228'. The conductive layer 238 is insulated from the floating gate 228' and the substrate 200 by the oxide layer 236.

20 In FIG. 14, an optional conductive layer 242, such as tungsten silicide, having a thickness of about 500 Å is formed on the conductive layer 238 by the same approach of patterning the conductive layer 238 to reduce the resistance of the conductive layer 238. Thereafter, an insulating spacer 244, such as silicon nitride, having a thickness of about 1300 Å is formed on the sidewall of the insulating stud 232' to cover a part of the conductive layer 242.

25 In FIG. 15, the conductive layer 242 and the underlying conductive layer 238 and oxide layer 236 are etched using the

30

insulating spacer 244 as a mask to expose the substrate 200. The remaining conductive layer 238' is used as a control gate, and the remaining oxide layer 236' is used as a gate oxide.

5 In FIG.16, a conformable insulating layer (not shown), such as silicon nitride, is formed on the substrate according to the FIG. 15 by conventional deposition, such as CVD. Subsequently, the insulating layer is etched by anisotropic etching, such as dry etching, to form an insulating spacer 250 on the sidewall of the control gate 238'. Thereafter, a doping
10 region D is formed in the exposed substrate 200 by ion implantation. Next, annealing is performed. Here, the doping region D is used as a drain region. Next, an insulating layer 252, such as silicon oxide or borophosphosilicate glass (BPSG), is formed on the drain region D. Subsequently, the insulating
15 layer 252 is polished by CMP using the insulating spacer 244 as a stop layer.

In FIG. 17, lithography and etching are performed on the insulating layer 252 to expose the drain region D.

20 In FIG.18, a conductive plug 254, such as doped polysilicon, is formed on the drain region D serving as a bit line contact. Thereafter, a cap layer 256, such as tetraethyl orthosilicate (TEOS) oxide, is formed on the conductive plug 254, insulating spacer 244, and the insulating stud 232'.

25 In FIG. 19, dual damascene process is used to form a bit line as follow. Lithography and etching are performed on the cap layer 256 to remove the portion of the cap layer 256 on the conductive plug 254. Next, a conductive layer 258, such as tungsten, is formed on the conductive plug 254 and the remaining cap layer 256. Finally, the conductive layer 258 is polished

by CMP to form a bit line 258 and finish the split gate flash memory cell fabrication of the present invention.

Also, FIG. 19 is a cross-section showing the structure of the split gate flash memory cell according to the present invention. The memory cell includes a substrate 200, such as
5 a silicon wafer, having a trench. A conductive stud 216', such as polysilicon, is disposed in the lower trench serving as a source line. The source line 216' is insulated from the substrate 200 by an oxide layer 226.

10 A source region S is formed in substrate 200 adjacent to the upper source line 216'. A conductive spacer 228' is disposed on the sidewall of the upper trench and protrudes from the top of the substrate 200 serving as a floating gate. The floating gate 228' is insulated from the substrate 200 by the gate oxide
15 226.

The floating gate 228' is insulated from the source line 216' by an insulating layer 222", such as HDP oxide. An insulating stud 232' is disposed on the insulating layer 222". A conductive layer 238', such as doped polysilicon, is disposed
20 on the substrate 200 of the outside floating gate 228' serving as a control gate. The control gate 238' is insulated from the floating gate 228' and the substrate 200 by gate oxide 236'.

An insulating spacer 244 is disposed on the sidewall of the insulating stud 232' and an insulating spacer 250 is disposed
25 on the sidewall of the control gate 238'. These insulating spacers 244, 250 can be silicon nitride. A conductive layer 242, such as tungsten silicide, is disposed between the insulating spacer 244 and the control gate 238'. A drain region D is formed in the substrate 200 of the outside control gate 238'.

A conductive plug 254, such as doped polysilicon, is disposed on the drain region D serving as a bit line contact. A cap layer 256, such as TEOS oxide, is disposed on the insulating stud 232' and the insulating spacer 244. A conductive layer 258, such as tungsten, is disposed on the conductive plug 254 and the cap layer 256 serving as a bit line.

Compared to the prior art, in the split gate flash memory cell of the invention, the floating gate and source line are disposed in the trench to effectively increase the integration of ICs by reducing the size of the memory cell. Moreover, no alignment issues occur during control gate fabrication, and the critical dimension (CD) can be controlled by adjusting the thickness of the insulating spacer over the control gate. In addition, the conductive plug is formed by self-alignment to simplify the process of fabricating the split gate flash memory.

The foregoing description has been presented for purposes of illustration and description. Obvious modifications or variations are possible in light of the above teaching. The embodiments were chosen and described to provide the best illustration of the principles of this invention and its practical application to thereby enable those skilled in the art to utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. All such modifications and variations are within the scope of the present invention as determined by the appended claims when interpreted in accordance with the breadth to which they are fairly, legally, and equitably entitled.